DOCUMENT RESUME

ED 218 341 TM 820 382

AUTHOR Percy, Virginia R.; Smith, Richard M.
TITLE The Use of Rasch Ability Estimates for English and Mathematics Placement.
PUB DATE Mar 82
NOTE 16p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New York, NY, March, 1982); For a related document, see ED 189 174.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; Basic Skills; Computer Oriented Programs; *Difficulty Level; *Goodness of Fit; *Item Analysis; *Latent Trait Theory; Screening Tests; *Student Placement; Test Items; Two Year Colleges
IDENTIFIERS New Jersey College Basic Skills Placement Test; *Rasch Model

ABSTRACT

Replicating a 1978 study of methods of setting cut-off scores in standardized tests for student placement, the Rasch model was used to estimate ability levels and item difficulty for use in determining the probability of course success. Community college students were administered the New Jersey College Basic Skills Placement Tests and items were calibrated by the BICAL program. Item difficulties were analyzed using a three-course mathematics sequence, a remedial mathematics course, and a three-course English sequence. Success was generally defined by earned grade. The comparisons of different grade cut-offs showed the course grades can be transformed into dichotomous success "items" which fit the Rasch model. (Auth/CM)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED218341

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

# THE USE OF RASCH ABILITY ESTIMATES FOR ENGLISH AND MATHEMATICS PLACEMENT

Virginia R. Percy
Arthur Andersen & Co.

Richard M. Smith
University of Chicago

March, 1982

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

V.R. Percy

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the annual meeting of the National Council on Measurement in Education.

New York - March, 1982

TM820382

INTRODUCTION

The concern among educators and the public about many students' needs for increased exposure to education in basic skills has grown over the past two decades as standardized test scores have dropped. Decisions made about student placement into remedial and other mathematics and English courses in post-secondary institutions have consequently become increasingly important.

Persistent problems have been connected with the use of standardized tests for the placement of students into mathematics and English courses. In particular, there have been no guidelines except past experience, subjective judgment and test norms for setting the cut-off scores. Existing methods for determining these scores (Nedelsky, 1954; Angoff, 1971; Ebel, 1972) have employed panels of judges to define such concepts as "minimum competency" and "mastery" within the context of a particular test. Different methods have been shown to yield widely differing results (Andrews and Hecht, 1976; Bernknopf, 1980; Kleinke, 1980).

This paper describes a study which was carried out to expand upon previous work in the area of cut-off scores. Both studies employed Rasch model calibrations of placement test items into which a "success item" had been inserted (Wright, 1977). This dichotomous "item" was defined in terms of course grades ("A," "B" or "C" = 1; below "C" = 0), and, because the Rasch model yields estimates of abilities and item difficulties in the same linear metric, its difficulty can be interpreted in terms of an individual's probability of success in the course. A person with an ability exactly equal to the difficulty of the success "item" would have a 50% chance of earning a grade of "C" or above. Not only can this probability be calculated using an individual's estimated ability, but an examination of the test items above and below the success "item" can tie it directly to the content being tested. The difficulty

of this item can then be seen as a possible cut-off score and also given a content interpretation. Other cut-off scores can also be established which correspond to probabilities of success above or below 50%.

## THE STUDY

The previous study (Mitchell and Smith, 1980) took a preliminary look at a three-course mathematics sequence offered at a New Jersey community college. The placement test in use was that which had been mandated for all public post-secondary institutions in the state, the New Jersey College Basic Skills Placement Test (BSPT) published by ETS.

The present study looked at the same sequence of mathematics courses (MA 109, MA 122 and MA 103) for the following year, i.e., 1979. One difference was that 20% of the BSPT items had been replaced as part of a continuing program of test security. In addition, the lowest level remedial math course (MS 100) was included, thus completing a four-course sequence. A new aspect of the study was a similar examination of the equivalent three-course English sequence. This consisted of a remedial course, ES 100, a composition course, EG 101, and a literature and composition course, EG 102. Finally, one of the math courses was examined in more detail by altering the definition of "success" from that of a grade of "C" or above to, first, earning an "A" or a "B," and, then, to earning only an "A." This effectively allowed the examination of different cut-off scores associated with earning different grades in the same course.

The goals for the present work, therefore, were to:

1. Determine how well the previous study was replicated.
2. Examine how well the remedial mathematics course fit into the pattern of the first three.
3. Examine the results for the three-course English sequence.
4. Compare the results for the two content areas.

5. Examine the various cut-offs that are associated with different grades in a single course.

PROCEDURES

The procedures for the previous study were again followed. The BICAL program (Wright and Mead, 1978) was used to calibrate the BSPT items. The standard deletions of respondents with perfect scores or scores below a chance level and items answered either correctly or incorrectly by everyone in the sample were made. Dichotomized success variables were created from students' grades in the courses. For all seven courses, success was defined as a grade of "C" or more. In addition, for the second mathematics course, MA 109, two additional calibrations were run with success being defined as a grade of "A" or "B" and a grade of "A," respectively. BICAL generated between and total fit statistics with an expected value of "1" for each item which with its associated standard error, indicated how well each item, including the success "items," fit the model.

The Computation sub-test alone had been used for student placement into the two lower mathematics courses while the sum of the scores from the Computation plus Elementary Algebra sub-tests was used for placement into the two higher courses. Separate calibrations were run for each course with the inclusion of the success "item" for that course. Although each course was made up of different students, they all took the Computation sub-test. Thus, the item difficulties from the calibrations for the first two courses could be directly compared, and, when adjusted by a constant, they could be compared to those for the other two courses. This constant was determined by calculating the mean for the difficulties for the Computation sub-test items for the second and third courses in the sequence. The difference in the two means was the constant of alignment and was subtracted from the difficulties of the lower level courses.

The same procedure was followed for the three-course English sequence. In this case, however, all placement decisions were made on the basis of a sum of scores from the Reading Comprehension, Sentence Structure and Logical Relationships sub-tests. The item difficulties and those for the success "items" could, therefore, be compared without any adjustment.

The final part of the study involved an in-depth examination of the second course in the mathematics sequence, MA 109, as different definitions were applied to the "success" item. The difficulty for the success "item" with the standard definition (a grade of "C" or higher) was determined above. Two further calibrations were run with success defined as, first, a grade of "B" or above and, then, as a grade of "A." In this way, the difficulties associated with earning an "A," a "B" or better, and a "C" or better could be compared for this one course and could also be compared with the difficulties of the success "items" for the other mathematics courses.

## RESULTS

Replication of Previous Work. Table I shows the results of the introduction of a success "item" into the BSPT calibrations from 1978 and 1979 for all but the lowest course of the four mathematics courses studied. During the time between the two studies, some changes were made to the section of BICAL which calculates the fit statistics. The former version calculated "f" statistics while the current one uses "t's." Thus, the expected value was "1" in the first study and "0" in the second. Any item with a "t" of more than two could not be said to fit the model.

The total "t"-statistics for the success "items" for the three math courses (MA 109, MA 122 and MA 103) were -.25, -1.08 and -.4, respectively. Thus, two were less than, and one only slightly more than, one standard deviation (expected value of 1) from "0." BICAL also generates a between-group "t"

TABLE I: COMPARISON OF TWO STUDIES

| | MA 109 | | MA 122 | | MA 103 | |
|---|---|---|---|---|---|---|
| | 1978 | 1979 | 1978 | 1979 | 1978 | 1979 |
| Number of Subjects | 122 | 149 | 128 | 86 | 72 | 33 |
| BSPT Subtests | Computation | | Computation<br>Elementary Algebra | | Computation<br>Elementary Algebra | |
| Number of Items | 30 | | 60 | | 60 | |
| "Success Item": Item Difficulty | .58[1] | -.69 | .58 | -.14 | .82 | .55 |
| Raw Score Equivalent | 37 | 22 | 37 | 29 | 40 | 37 |
| Standard Error | .21[1] | .19 | .19 | .24 | .26 | .39 |
| Total Fit | 1.05[2] | -.25[3] | .9 | -1.08 | 1.04 | -.4 |
| Between-group Fit | .55[2] | .26[3] | .43 | .43 | 1.41 | 2.00 |

1. Units are logits
2. "f" statistics
3. "t" statistics

statistic to analyze the fit of items across six user-specified subgroups. These subgroups were defined as low-to-high ability groupings for each course. For the three math courses the between-group "t's" were .26, .43 and 2.00, respectively. The first two are quite satisfactory, but the third is barely acceptable. An examination of the percent of respondents who were successful in the highest course, MA 103, across the six subgroups indicated that not as many in the lowest ability group were as successful as the model had predicted. This was the cause of the relative lack of fit. The between-group "t" of 2.00 was still less, however, than the between fit statistics calculated for four other items from the BSPT.

It can be seen that the success "items" fit the model about as well as did those from the previous year. What is different, however, is the difficulties of these "items." The first study showed that success in the two lower courses was essentially equally difficult to achieve. That is, the first of the three courses in the hierarchical sequence was as difficult as the second. Also, the difficulty for the success "item" for the third course was only .24 logits, or about one standard error, above the first two.

One reason why the first math course in this sequence (MA 109) was so difficult was that it was designed in a "Mastery" format. Students were required to pass 15 weekly tests with a score of 80% or above and could not advance to the next test until all previous ones had been passed. They were allowed to retake parallel exams of every test several times. Students could only earn a grade of "A" or "B" in the course, and over two thirds failed every semester.

Attempts were made to improve this situation, and during the time between the first and second studies, standards were lowered somewhat, allowing students to earn grades of "C." It can be seen that the difficulty of the success "item" for this course decreased 1.27 logits. After this change, in

fact, success in all three courses became easier, and the difficulties for success in the second and third courses decreased .72 and .27 logits repectively (see Figure 1 below). Thus, while the difficulties were within one standard error of each other for all three courses in the first study, the success "item" for each course was at least 1½ to 2 standard errors from that of the next course in the sequence in the second study. The original 3-point raw score range had expanded to a 15-point range.

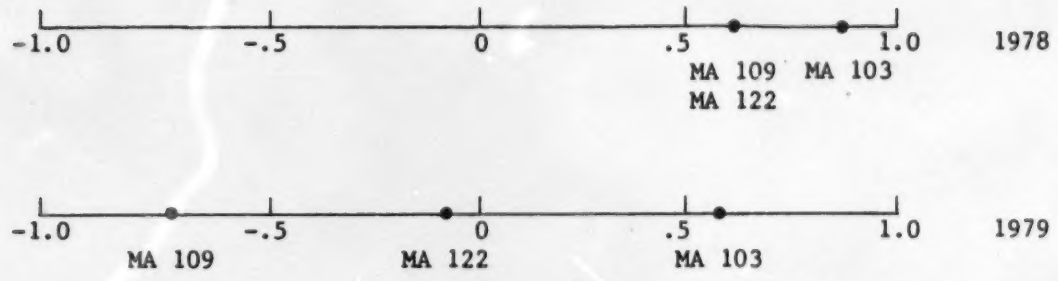


Figure 1

Difficulties of 1978 and 1979 Mathematics
Success "Items" (in Logits)

Finally, the constant of adjustment was .78 logits this year, as compared with .75 logits last year. Even though the 1979 version of the BSPT contained only 80% of the same terms as did the previous year's test, the "Computation" and "Elementary Algebra" sub-tests behaved similarly for both years.

Fit of the Remedial Mathematics Course. The lowest level remedial math course was added to the second study, and Table II indicates that the difficulty of its success "item" was 1.28 logits below that of the first course in the three-course sequence. It has a standard error of .19 logits and fits the model as well as the other three success "items." Thus, the

TABLE II

SUCCESS "ITEMS" FOR FOUR MATHEMATICS COURSES

| Course | No. of Students | Item Difficulty (Logits) | Raw Score Equivalents | Standard Error (Logits) | Total "t" | Between-Group "t" |
|---|---|---|---|---|---|---|
| MS100 | 136 | -1.97 | 10 | .19 | -1.12 | .68 |
| MA109 | 149 | - .69 | 22 | .19 | - .25 | .26 |
| MA122 | 86 | - .14 | 29 | .24 | -1.08 | .43 |
| MA103 | 33 | .55 | 37 | .39 | - .4 | 2.00 |

TABLE III

SUCCESS "ITEMS" FOR THREE ENGLISH COURSES

| Course | No. of Students | Item Difficulty (Logits) | Raw Score Equivalents | Standard Error (Logits) | Total "t" | Between-Group "t" |
|---|---|---|---|---|---|---|
| ES100 | 194 | - .67 | 45 | .16 | 1.06 | .34 |
| EG101 | 285 | .49 | 76 | .13 | - .8 | .15 |
| EG102 | 129 | 1.38 | 95 | .21 | 1.35 | 1.49 |

TABLE IV

SUMMARY OF INFORMATION ON MA 109 GRADES

| Success Definition | No. of Students | Difficulty of Success "Item" | Standard Error | Total "t" | Between-Group "t" |
|---|---|---|---|---|---|
| A | 154 | 1.46 | .22 | .99 | .38 |
| A or B | 152 | .13 | .18 | .15 | 1.47 |
| A, B, or C | 149 | -.69 | .19 | -.25 | .26 |

range in difficulties for success from the lowest to the highest of these four courses is 2.52 logits or the equivalent of 27 raw score points out of 60.

Results for the English Sequence. As can be seen from Table III, the three courses in the English sequence have success "items" with difficulties which are clearly distinct. The first course, ES 100, Remedial English, and the second course, EG 101, College-level Composition, were 1.16 logits apart. The second course and the third, EG 102, literature, were .89 logits apart. This spanned roughly 50 raw score points on the 125 items in the combined score from the Reading Comprehension, Sentence Structure, and Logical Relationships Sub-tests of the BSPT which were used for English placement.

The success "items" for the first two English courses clearly fit the model, falling roughly one standard deviation or less from the expected value of "0." The success "item" for the third course still fit the model but not as well as for the other two courses with a between-group fit of 1.49 logits. When the percents correct for the six sequential ability groups were compared, the source of this lack of fit becomes clear. The second ability group, which had a mean ability more than a logit higher than the lowest group, had a lower percent correct (or rate of "success") than the latter. There were nine items on the three BSPT sub-tests with higher between-group fit statistics, however.

Thus, the success "items" for the three-course English sequence behaved as they were expected to, i.e., success in the lowest course required the least ability, and so on. They all fit the model acceptably.

Comparing the Two Subject Areas. It can be seen, then, that abilities associated with success in both subject areas have the same hierarchical patterns as do the two sets of courses. Both the English and mathematics variables are fairly well-defined as indicated by the degree of separation of these abil-

ities from course to course. The English sequence forms a more definite hierarchy as there are 1.16 logits between the difficulties of the first and second success "items" and .89 between those of the second and third. This is equivalent to more than four standard errors in each case.

The math sequence, while not as well-defined, is certainly more well-defined for this study than it was in the previous one. Success in the Elementary Algebra course (MA 109), as described above, most likely resulted originally from perseverance as much as ability. As efforts were taken to make this course easier, the four-course math hierarchy became more reflective of math ability alone and, hence, more clearly defined.

Different Grade Cut-Offs. In addition to the definition of success used throughout this paper, i.e., a grade of "C" or above, two other BICAL calibrations were run using two stricter definitions for the Elementary Algebra course (MA 109). Thus, success "items" associated with earning a grade of "B" or above or a grade of "A" could be compared with the original "item."

Table 4 shows the hierarchy of abilities which these "items" represent. There are several interesting things about it. First, these "items" are as well separated as are those for the content sequences discussed above. They also fit the model equally well. The ability associated with earning a grade of "B" or above (.13 logits) is greater than that required to succeed with a "C" or above in the next highest course, MA 122. Also, earning an "A" in MA 109 requires a much higher ability (1.46 logits) than that required to earn a "C" or above in the highest course, MA 103. These relationships are shown in Figure 2 on the next page.

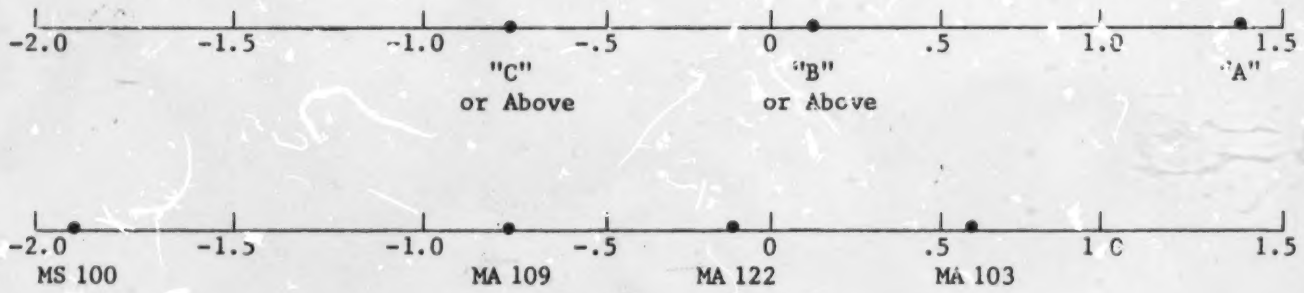


Figure 2

1979 Mathematics Success "Items"
and MA 109 Success/Grade Levels (in Logits)

Secondly, when MA 109 was offered at the time of the first study, as discussed above, only grades of "A" and "B" were given. Thus, when the success "item" from this study is compared with the next year's success "item" for a grade of "B" or above, it can be seen that it did become easier to earn the same grade in the same course in 1979 (.13 logits) than it had been in 1978 (.58 logits).

Finally, an examination of the items on the BSPT with difficulties which fall around the three identified cut-offs can produce a content interpretation for this course, i.e., how the various grade levels translate into abilities to perform various mathematical tasks. Table 5 illustrates this content-ability match, thereby defining the mathematics variable for one course in this study. It can be seen that an item is difficult if it calls for a higher level skill but also if it requires knowledge of less commonly-used notation. Also, it is known that word problems are difficult, and Table 5 shows that word problems with percents and fractions are more difficult than those which only ask for a computation of an average.

## IMPORTANCE OF THE STUDY

This study has replicated the previous one by again showing that course

Definition of the Mathematics Variable for MA 109

TABLE 5

| Task | Difficulty (Logits) | Example |
|---|---|---|
| Earning a grade of "A" | 1.46 | |
| Division by fractions with slash notation | 1.2 | 8/¼ |
| Percent problem | .60 | 13 is 6% of ? |
| Word problems with percent or fractions | .22 to .32 | A recipe calls for two cups of sugar for every cup of milk. If two cups of milk are used, how many cups of sugar should be used? |
| Earning a grade of "B" or above | .13 | |
| Division by fractions with "÷" notation | -.46 | $5 \div \frac{3}{20}$ |
| Earning a grade of "C" or above | -.69 | |
| Subtraction or multiplication of mixed numbers | -.70 | $\begin{array}{r} 6\ 1/2 \\ -\ 3\ 2/3 \\ \hline \end{array}$ |
| Subtraction of fractions | -1.0 | 3/4 - 1/12 |
| Word problem calling for computation of an average | -1.2 | The weights of four children are 70, 84, 75, and 90 pounds respectively. What is their average weight in pounds? |
| Subtraction of whole numbers with carrying | -2.8 | $\begin{array}{r} 5006 \\ -\ 3289 \\ \hline \end{array}$ |

grades can be transformed into dichotomous success "items" which fit the Rasch model. These "items" can then be placed on a continuum with other items which define the variable and have the advantage of being "sample-free." They can be further defined in terms of the associated tasks which can be accomplished at equivalent ability levels.

The examination of the shifts in difficulties of the success "items" for the mathematics sequence and, in particular, for the Elementary Algebra course (MA 109), indicated how changes in grading and administrative policies affecting student success also affect the success "items." Thus, the quality of cut-off scores derived from such "items" can be continually monitored and these scores may be adjusted on the basis of empirical data, rather than by hunch. Continued study of these "items" across a course sequence can provide the information needed to know when the hierarchy is not functioning as such and when one or more courses are in need of more in-depth evaluation.

Finally, this study and the previous one suggest that decisions concerning placement into English and mathematics courses may be made with a good deal more data than were previously available. Such data would include the probability of success in a given course or of earning a particular grade in that course based upon a student's ability as estimated from his or her placement test score. It may well become feasible to make such information public and to involve the students in deciding which course or combination of courses would combine required subject areas so as to maximize their probability of success.

## References

Andrews, B. J. and Hecht, J. T., A preliminary investigation of two procedures for setting examination standards, Educational and Psychological Measurement, 1976, 36, 45-50.

Angoff, W. H., Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), Educational Measurement (2nd edition). Washington, D. C.: American Council on Education, 1971.

Bernknopf, S., Cut-scores and alternate forms: A new frontier or back to the trenches. Paper presented to the annual meeting of the National Council on Measurement in Education, Boston, 1980.

Ebel, R. L., Essentials of Educational Measurement. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Hambleton, R. K. and Cook, L. L., Latent Trait Models and Their Use in the Analysis of Educational Test Data. Journal of Educational Measurement, 1977, vol. 14, no. 2, 75-96.

Hambleton, R. K., Swaminathan, H., Cook, L. L., Eignor, D. R., and Gifford, J. A., Developments in latent trait theory: Models, technical issues, and applications. Review of Educational Research, 1978, vol. 48, 4, 467-510.

Kleinke, D. J., Applying the Angoff and Nedelsky techniques to the National Licensing Examinations in Landscape Architecture. Paper presented to the annual meeting of the National Council on Measurement in Education, Boston, 1980.

Mitchell, V. P. and Smith, R. M., Determining cut-off scores using Rasch ability estimates. Paper presented to the annual conference of the National Council on Measurement in Education, Boston, 1980.

Nedelsky, L., Absolute grading standards for objective tests. Educational and Psychological Measurement, 1954, 14, 3-19.

Smith, R. and Mitchell, V. P., Improving the predictive ability of placement tests using the Rasch model for scoring. Paper presented to the annual conference of the National Council on Measurement in Education, San Francisco, 1979.

Waller, M., Ability removing random guessing model. Paper presented at the annual conference of the American Educational Research Association, San Francisco, 1976.

Wright, B. D., Solving measurement problems with the Rasch model. Journal of Educational Measurement. 1977, vol. 14, no. 2, 97-116.

Wright, B. D. and Mead, R. J., BICAL: Calibrating items and scales with the Rasch Model. Research Memorandum Number 23A, University of Chicago, 1978.